Kyle Welch

IST 718

Lab 1

**Data**

The data used in this lab begins with the Coaches (‘Coaches9.csv’) which was provided as part of the lab. To further expand and ‘widen’ the data the following sources were used: For graduation rates (GSR and FGR) <https://web3.ncaa.org/aprsearch/gsrsearch> was utilized. For all time win records, 2019-win records, and stadium capacity [www.sports-reference.com](http://www.sports-reference.com) was utilized. Finally, donation data was gathered from <http://cafidatabase.knightcommission.org/fbs>.

**Cleaning**

The data was pulled manually from each of the corresponding websites and the put into a CSV file to then be loaded into Google Colab. All five of the data sets, including Coaches, was loaded into a pandas data frame. Following which a column that appeared in multiple of the data sets, ‘Conferences’, was dropped due to redundancy. Next, columns that were determined to be unnecessary for this lab were dropped from the data frame, including TotalPay, Bonus, BonusPay, Buyout, AssistantPay, below is the reasoning of why they were dropped.

* TotalPay
  + SchoolPay was kept in favor of TotalPay, TotalPay was dropped due to redundancy.
* Bonus & BonusPay
  + Both were dropped due to inconsistencies of how the data was provided, there would be a 0 or a blank in one column and an actual figure in the other. Additionally, there were no plans to use either in the regression analysis.
* Buyout
  + Buyout was dropped as it was determined to be not useful for the regression analysis in this case.
* AssistantPay
  + AssitantPay was dropped as all values were 0.

All rows with any NaNs from the remaining columns were dropped, in the question section a more in-depth breakdown of which schools were dropped will be provided. The last two data cleaning process that took place was the data type of SchoolPay was changed from an object to a float64 and three columns that were introduced names were changed to prevent problems in the analysis.

**Analysis**

**Icon

Description automatically generated**

The above density plot was created to visualize the current coaches’ salary across the FBS schools. This visualization should provide a good idea of what a likely salary is for any FBS coach before the salary of specifically the Syracuse coach is considered.

Chart, histogram

Description automatically generated

The heatmap on the previous page represents the correlation between variables, providing a view of which variables have the greatest impact on SchoolPay, as well as one another.

**Regression**

At first, all variables were considered; however, four models were generated using a slightly different combination each time. Each regression model will be noted with which variables were not considered. Additionally, all regression models will be followed by the salary the model recommends for the Syracuse head coach.

All variables (none dropped)

OLS Regression Results

==============================================================================

Dep. Variable: SchoolPay R-squared: 0.855

Model: OLS Adj. R-squared: 0.822

Method: Least Squares F-statistic: 25.72

Date: Sun, 31 Jan 2021 Prob (F-statistic): 2.58e-24

Time: 02:15:39 Log-Likelihood: -1377.6

No. Observations: 92 AIC: 2791.

Df Residuals: 74 BIC: 2837.

Df Model: 17

Covariance Type: nonrobust

==========================================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------------------

Intercept -2.921e+06 1.27e+06 -2.295 0.025 -5.46e+06 -3.85e+05

Conference[T.ACC] 8.84e+05 5.88e+05 1.503 0.137 -2.88e+05 2.06e+06

Conference[T.Big 12] 8.074e+05 5.9e+05 1.368 0.176 -3.69e+05 1.98e+06

Conference[T.Big Ten] 1.117e+06 5.46e+05 2.045 0.044 2.85e+04 2.21e+06

Conference[T.C-USA] -3.93e+05 5.55e+05 -0.708 0.481 -1.5e+06 7.12e+05

Conference[T.Ind.] -1.879e+05 9.89e+05 -0.190 0.850 -2.16e+06 1.78e+06

Conference[T.MAC] -3.95e+05 5.33e+05 -0.740 0.461 -1.46e+06 6.68e+05

Conference[T.Mt. West] -2.896e+05 5.38e+05 -0.538 0.592 -1.36e+06 7.83e+05

Conference[T.Pac-12] 1.793e+05 5.28e+05 0.339 0.735 -8.74e+05 1.23e+06

Conference[T.SEC] 8.325e+05 6.1e+05 1.366 0.176 -3.82e+05 2.05e+06

Conference[T.Sun Belt] -4.688e+05 5.69e+05 -0.824 0.413 -1.6e+06 6.65e+05

GSR 2.399e+04 1.77e+04 1.355 0.180 -1.13e+04 5.93e+04

FGR -6308.3761 1.39e+04 -0.454 0.651 -3.4e+04 2.14e+04

TotalGames -136.1511 451.195 -0.302 0.764 -1035.177 762.874

ATWinPercentage 2.947e+06 1.45e+06 2.028 0.046 5.18e+04 5.84e+06

WinPercentage19 6.051e+05 4.99e+05 1.212 0.229 -3.89e+05 1.6e+06

Capacity 26.2524 8.888 2.954 0.004 8.542 43.963

AnnualDonations 0.0315 0.011 2.881 0.005 0.010 0.053

==============================================================================

Omnibus: 7.726 Durbin-Watson: 2.115

Prob(Omnibus): 0.021 Jarque-Bera (JB): 9.816

Skew: 0.394 Prob(JB): 0.00739

Kurtosis: 4.393 Cond. No. 4.65e+08

Recommended salary for Syracuse coach: $5,514,519

Recommended salary for Syracuse coach, if Syracuse moved to Big Ten: $5,747,456

FGR, Total Games Played, and 2019 Win Percentage were dropped for this regression

OLS Regression Results

==============================================================================

Dep. Variable: SchoolPay R-squared: 0.851

Model: OLS Adj. R-squared: 0.824

Method: Least Squares F-statistic: 31.53

Date: Sun, 31 Jan 2021 Prob (F-statistic): 3.86e-26

Time: 02:22:48 Log-Likelihood: -1378.7

No. Observations: 92 AIC: 2787.

Df Residuals: 77 BIC: 2825.

Df Model: 14

Covariance Type: nonrobust

==========================================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------------------

Intercept -3.222e+06 1.23e+06 -2.627 0.010 -5.66e+06 -7.8e+05

Conference[T.ACC] 7.269e+05 5.69e+05 1.278 0.205 -4.06e+05 1.86e+06

Conference[T.Big 12] 6.567e+05 5.63e+05 1.166 0.247 -4.65e+05 1.78e+06

Conference[T.Big Ten] 9.737e+05 5.23e+05 1.862 0.066 -6.76e+04 2.01e+06

Conference[T.C-USA] -3.105e+05 5.3e+05 -0.585 0.560 -1.37e+06 7.46e+05

Conference[T.Ind.] -3.404e+05 9.64e+05 -0.353 0.725 -2.26e+06 1.58e+06

Conference[T.MAC] -4.393e+05 5.28e+05 -0.831 0.408 -1.49e+06 6.13e+05

Conference[T.Mt. West] -3.003e+05 5.32e+05 -0.565 0.574 -1.36e+06 7.58e+05

Conference[T.Pac-12] 8.903e+04 5.14e+05 0.173 0.863 -9.35e+05 1.11e+06

Conference[T.SEC] 7.111e+05 5.77e+05 1.232 0.222 -4.38e+05 1.86e+06

Conference[T.Sun Belt] -4.448e+05 5.41e+05 -0.822 0.414 -1.52e+06 6.32e+05

GSR 2.205e+04 1.14e+04 1.933 0.057 -664.304 4.48e+04

ATWinPercentage 3.316e+06 1.38e+06 2.402 0.019 5.67e+05 6.07e+06

Capacity 27.3222 8.776 3.113 0.003 9.848 44.797

AnnualDonations 0.0345 0.011 3.283 0.002 0.014 0.055

==============================================================================

Omnibus: 10.268 Durbin-Watson: 2.126

Prob(Omnibus): 0.006 Jarque-Bera (JB): 12.506

Skew: 0.569 Prob(JB): 0.00192

Kurtosis: 4.402 Cond. No. 4.58e+08

Recommended salary for Syracuse coach: $ 5,864,615

Recommended salary for Syracuse coach, if Syracuse moved to Big Ten: $ 6,111,471

Both graduation rates (GSR & FGR) were dropped for this regression

OLS Regression Results

==============================================================================

Dep. Variable: SchoolPay R-squared: 0.850

Model: OLS Adj. R-squared: 0.821

Method: Least Squares F-statistic: 28.81

Date: Sun, 31 Jan 2021 Prob (F-statistic): 2.87e-25

Time: 02:24:58 Log-Likelihood: -1379.1

No. Observations: 92 AIC: 2790.

Df Residuals: 76 BIC: 2830.

Df Model: 15

Covariance Type: nonrobust

==========================================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------------------

Intercept -1.441e+06 8.29e+05 -1.738 0.086 -3.09e+06 2.1e+05

Conference[T.ACC] 9.513e+05 5.85e+05 1.627 0.108 -2.13e+05 2.12e+06

Conference[T.Big 12] 8.395e+05 5.92e+05 1.419 0.160 -3.39e+05 2.02e+06

Conference[T.Big Ten] 1.243e+06 5.41e+05 2.297 0.024 1.65e+05 2.32e+06

Conference[T.C-USA] -4.362e+05 5.53e+05 -0.789 0.432 -1.54e+06 6.64e+05

Conference[T.Ind.] -1.21e+05 9.86e+05 -0.123 0.903 -2.09e+06 1.84e+06

Conference[T.MAC] -3.821e+05 5.35e+05 -0.714 0.477 -1.45e+06 6.83e+05

Conference[T.Mt. West] -2.715e+05 5.38e+05 -0.505 0.615 -1.34e+06 7.99e+05

Conference[T.Pac-12] 2.615e+05 5.27e+05 0.496 0.621 -7.88e+05 1.31e+06

Conference[T.SEC] 9.467e+05 6.04e+05 1.568 0.121 -2.56e+05 2.15e+06

Conference[T.Sun Belt] -4.909e+05 5.71e+05 -0.860 0.392 -1.63e+06 6.46e+05

TotalGames -62.0735 449.930 -0.138 0.891 -958.186 834.039

ATWinPercentage 2.593e+06 1.44e+06 1.806 0.075 -2.67e+05 5.45e+06

WinPercentage19 8.427e+05 4.76e+05 1.771 0.081 -1.05e+05 1.79e+06

Capacity 27.1225 8.875 3.056 0.003 9.446 44.799

AnnualDonations 0.0278 0.011 2.599 0.011 0.006 0.049

==============================================================================

Omnibus: 6.667 Durbin-Watson: 2.129

Prob(Omnibus): 0.036 Jarque-Bera (JB): 8.967

Skew: 0.288 Prob(JB): 0.0113

Kurtosis: 4.417 Cond. No. 4.29e+08

Recommended salary for Syracuse coach: $ 4,010,301

Recommended salary for Syracuse coach, if Syracuse moved to Big Ten: $ 4,302,178

Both graduation rates, Total Games Played, and All-Time Win Percentage were dropped for this regression

OLS Regression Results

==============================================================================

Dep. Variable: SchoolPay R-squared: 0.844

Model: OLS Adj. R-squared: 0.818

Method: Least Squares F-statistic: 32.46

Date: Sun, 31 Jan 2021 Prob (F-statistic): 4.18e-26

Time: 02:28:41 Log-Likelihood: -1381.0

No. Observations: 92 AIC: 2790.

Df Residuals: 78 BIC: 2825.

Df Model: 13

Covariance Type: nonrobust

==========================================================================================

coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------------------

Intercept -5.926e+05 5.85e+05 -1.013 0.314 -1.76e+06 5.72e+05

Conference[T.ACC] 9.716e+05 5.81e+05 1.673 0.098 -1.84e+05 2.13e+06

Conference[T.Big 12] 7.999e+05 5.8e+05 1.378 0.172 -3.56e+05 1.96e+06

Conference[T.Big Ten] 1.217e+06 5.3e+05 2.296 0.024 1.62e+05 2.27e+06

Conference[T.C-USA] -3.054e+05 5.34e+05 -0.572 0.569 -1.37e+06 7.57e+05

Conference[T.Ind.] -2.172e+05 9.9e+05 -0.219 0.827 -2.19e+06 1.75e+06

Conference[T.MAC] -2.446e+05 5.34e+05 -0.458 0.648 -1.31e+06 8.18e+05

Conference[T.Mt. West] -1.109e+05 5.34e+05 -0.208 0.836 -1.17e+06 9.53e+05

Conference[T.Pac-12] 3.086e+05 5.24e+05 0.589 0.558 -7.35e+05 1.35e+06

Conference[T.SEC] 8.516e+05 5.9e+05 1.444 0.153 -3.23e+05 2.03e+06

Conference[T.Sun Belt] -2.924e+05 5.42e+05 -0.540 0.591 -1.37e+06 7.86e+05

WinPercentage19 1.01e+06 4.68e+05 2.158 0.034 7.8e+04 1.94e+06

Capacity 33.5815 8.189 4.101 0.000 17.279 49.884

AnnualDonations 0.0295 0.011 2.772 0.007 0.008 0.051

==============================================================================

Omnibus: 8.853 Durbin-Watson: 2.087

Prob(Omnibus): 0.012 Jarque-Bera (JB): 14.804

Skew: 0.322 Prob(JB): 0.000610

Kurtosis: 4.857 Cond. No. 4.18e+08

Recommended salary for Syracuse coach: $ 3,067,088

Recommended salary for Syracuse coach, if Syracuse moved to Big Ten: $ 3,312,191

**Questions**

* **What is the recommended salary for the Syracuse football coach?**   
  Based on the four models created Syracuse’s recommended football coach salary varies from $3,067,088 to $5,864,615. Syracuse’s current head coach, Dino Babers, annual salary is $2,401,206 for reference.
* **What if we went to the Big Ten?**Despite the four different models, Big Ten coaches are roughly paid about $300,000 more, so it would be expected for Syracuse’s Big Ten coach salary to be in the $3,312,191 to $6,111,471 range.
* **What schools did we drop from our data, and why?**The total number of schools in the data dropped from 129 to 92, for a drop of 37 schools. Below is a list of schools that were dropped because Annual Donation data was unavailable.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Air Force | Duke | Miami (Ohio) | Notre Dame | Stanford | Vanderbilt |
| Army | Florida International | Mississippi | Pittsburgh | Syracuse | Wake Forest |
| Baylor | Hawaii | Navy | Rice | Texas Christian |  |
| Boston College | Liberty | Nevada-Las Vegas | South Florida | Texas-El Paso |  |
| Brigham Young | Louisiana-Lafayette | North Carolina State | Southern California | Texas-San Antonio |  |
| Central Florida | Massachusetts | Northern Illinois | Southern Methodist | Tulane |  |
| Connecticut | Miami (Fla.) | Northwestern | Southern Mississippi | Tulsa |  |

* **What effect does graduation rate have on the projected salary?**GSR and FGR, the graduation rates, were the two lowest correlated variables with SchoolPay, FGR being the only variable that was negatively correlated with it. Between the two, GSR appears much for useful in the models that were created, though both ultimately were not considered in multiple of the models.
* **How good is our model? What is the single biggest impact on salary size?**All four models have very high R-squared and Adjusted R-squared’s indicating a strong goodness of fit for the regression models. With that said, all of the models are projecting the salary of the Syracuse head coach to be higher than what it is, from half a million to two and half a million greater than actual. This means one of three things, the models are not doing a good job of predicting the Syracuse head coach salary, Dino Barbers was signed to a bargain deal by Syracuse, or Dino Barbers is not a ‘Syracuse head coach’, meaning that when he was hired Syracuse did not intend to hire a coach of typical ‘Syracuse/ACC quality’. Ultimately, even if either of the latter two are the case the model can be tweaked further and even rethought to come up with a better product. Currently, the variable with the single biggest impact on salary size must be the conference the coach is signing with.

**Conclusion**

The models used do a seemingly decent job of creating a regression that can predict coaches’ salaries; however, the models consistently over compensate the Syracuse head coach and needs refinement. In terms of graduation rates, GSR should be the preferred metric over FGR, and for win percentage the past season also does a better job of predicting than the all-time winning percentage. Stadium capacity and annual donations to the program were also both extremely useful for these models. Finally, a variable that was not used but could be useful if the models were to be further tweaked, is year the coaches’ contract was agreed to, this would provide insight into whether more recent contracts are compensated to a greater extent than older contracts.